Journal of Structural Biology 190 (2015) 11–20

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

NMR structure and dynamics of Q4D059, a kinetoplastid-specific and conserved protein from *Trypanosoma cruzi*

Aracelys López-Castilla [a], Tirso Pons [b], José R. Pires [a,*]

[a] Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, RJ, Brazil
[b] Programa de Biología Estructural y Biocomputación, Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain

ARTICLE INFO

Article history:
Received 25 October 2014
Received in revised form 23 February 2015
Accepted 24 February 2015
Available online 5 March 2015

Keywords:
NMR
Protein structure
Kinetoplastids
Trypanosoma cruzi
Neglected diseases
Drug target

ABSTRACT

Q4D059 (UniProt accession number), is an 86-residue protein from *Trypanosoma cruzi*, conserved in the related kinetoplastid parasites *Trypanosoma brucei* and *Leishmania major*. These pathogens are the causal agents of the neglected diseases: Chagas, sleeping sickness and leishmaniases respectively and had recently their genomes sequenced. Q4D059 shows low sequence similarity with mammal proteins and because of its essentiality demonstrated in *T. brucei*, it is a potential target for anti-parasitic drugs. The 11 hypothetical proteins homologous to Q4D059 are all uncharacterized proteins of unknown function. Here, the solution structure of Q4D059 was solved by NMR and its backbone dynamics was characterized by ^{15}N relaxation parameters. The structure is composed by a parallel/anti-parallel three-stranded β-sheet packed against four helical regions. The structure is well defined by ca. 9 NOEs per residue and a backbone rmsd of 0.50 ± 0.05 Å for the representative ensemble of 20 lowest-energy structures. The structure is overall rigid except for N-terminal residues A^{9} to D^{11} at the beginning of β1, K^{38}, V^{39} at the end of helix H3 with rapid motion in the ps–ns timescale and G^{25} (helix H2), I^{68} (β2) and V^{78} (loop 3) undergoing internal motion in the μs–ms timescale. Limited structural similarities were found in protein structures deposited in the PDB, therefore functional inferences based on protein structure information are not clear. Q4D059 adopts a α/β fold that is slightly similar to the ATPase sub-domain IIB of the heat-shock protein 70 (HSP70) and to the N-terminal domain of the ribosomal protein L11.

1. Introduction

The unicellular kinetoplastid protozoa *Trypanosoma cruzi* is the etiological agent of Chagas disease, a human illness endemic in Central and South America, also known as American trypanosomiasis (Hotez et al., 2008; Stuart et al., 2008). It affects approximately 8–11 million people, mostly those living in poor areas, thus low priority is given for investments in the pharmaceutical industry and in public health programs dealing with this disease (Hotez et al., 2008). It annually causes around 14,000 deaths (Muñoz-Saravia et al., 2012; Rassi et al., 2010). The acute phase of the disease is often asymptomatic and not diagnosed, therefore around one third of the patients evolve to a chronic and incurable form of the illness, manifesting cardiac and/or digestive disorders (mega-colon, mega-esophagus) (Rassi et al., 2010). Heart failure in people infected by *T. cruzi* is among the main reasons of morbidity and mortality in endemic areas (Muñoz-Saravia et al., 2012).

Despite the huge social and economic burden associated with Chagas disease, no vaccines are currently available and drugs for treatment are inefficient, toxic, with long and expensive treatment schedules. To date, nifurtimox and benznidazole are the only approved drugs in use and in both cases parasite resistance have been reported (Bastos et al., 2013). In this scenario, the identification of new targets and the development of novel molecules for preventing and treating the disease are highly necessary.

The *T. cruzi* genome has been sequenced (El-Sayed et al., 2005), as well as the genomes of two other related parasites, *Trypanosoma brucei* (Berriman et al., 2005) and *Leishmania major* (Ivens et al., 2005), causal agents of African sleeping sickness and leishmaniasis, respectively. In each case, ca. 10,000 genes per haploid genome were identified and approximately 50% of them code for hypothetical proteins (HPs) of unknown function (Berriman et al., 2005; El-Sayed et al., 2005; Ivens et al., 2005). Several of these proteins are conserved in the three genomes and exhibit low sequence similarity with proteins from other organisms, an appropriate

* Corresponding author at: Av. Carlos Chagas Filho, 373 – Bloco E, sala 10, 21941-902 Rio de Janeiro, RJ, Brazil.
E-mail address: jrmpires@cnrmn.bioqmed.ufrj.br (J.R. Pires).

http://dx.doi.org/10.1016/j.jsb.2015.02.007
1047-8477/

feature of potential targets for anti-trypanosomal drugs. In addition, proteomic studies on *T. cruzi* have corroborated the expression of more than 2000 of those predicted HPs (Atwood et al., 2005).

Studies on the trypanosomatid-specific HPs can increase the knowledge about the cellular biology and pathogenesis of these parasites, providing new perspectives for drug development. Through a structural genomic approach, HP structure determination may contribute to infer its function in those cases where structural similarities to proteins with known function are found. In addition, NMR and structural data is one pre-requisite in the structure-based design and ligand screening.

Using bioinformatic tools at the TritrypDB database (http://tritrypdb.org/tritrypdb/) (Aslett et al., 2010), a conserved kinetoplastid-specific protein of unknown function was selected for NMR structural studies. The gene TcCLB.510347.29 codifies the protein Q4D059 (UniProt accession number) from *T. cruzi*, which is an 86 residues, approximately 10 kDa protein, without significant sequence identity to mammal proteins. Q4D059 protein expression was corroborated by proteomic analysis (Atwood et al., 2005). Additionally, its mRNA was detected during the epimastigote stage of *T. cruzi* by microarray studies of gene expression (Minning et al., 2009). Moreover, its orthologue in *T. brucei* (Tb09.160.0465) was shown to be essential in both bloodstream, and procyclic forms of the parasite, as well as in differentiation of procyclic to bloodstream form, by means of RNA interference studies (Alsford et al., 2011). All these features make Q4D059 a potential target for the development of novel anti-trypanosomal drugs. Thus NMR structural characterization can facilitate its use in further research.

Here we report, the NMR structure in solution of Q4D059, its backbone dynamics and structural homology analysis.

2. Material and methods

2.1. Protein samples

The gene for Q4D059 from *T. cruzi* was synthesized by GenScript (USA) and cloned into the plasmid pGEX 4T2 (GE Healthcare) for its expression as a glutathione-S-transferase (GST) fusion protein, localized at the N-terminus and cleavable by thrombin. The construct was transformed into *E. coli* BL21 (DE3) (Invitrogen) and cells were grown in LB medium at 37 °C to an OD_{600} between 0.6 and 0.8, followed by addition of 1 mM isopropyl β-D-thiogalactoside. The protein expression was at 37 °C, overnight. Cell pellets were re-suspended in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) containing 1X protease inhibitor cocktail (Sigma Aldrich) and 1 mM EDTA, cooled on a ice bath and lysed by sonication. Cell debris was harvested by centrifugation at 15,000 g for 45 min, at 4 °C. The supernatant was applied into a Glutathione-Sepharose column (GE Healthcare) for isolating the fusion protein. GST-fusion protein was eluted with 10 mM reduced glutathione (50 mM Tris–HCl, pH 8.0) and the GST tag cleaved by thrombin (Sigma–Aldrich) (1 unit per milligram of fusion protein, 16 h, 25 °C). Complete cleavage was checked by SDS–PAGE. After cleavage, Q4D059 protein was further purified by gel filtration on a Superdex 75 column (GE Healthcare) using PBS as running buffer. Fractions containing Q4D059 were concentrated using a 3 kDa-cutoff centrifugal filter device (GE Healthcare) and the sample buffer was exchanged to the NMR buffer (20 mM phosphate, 50 mM NaCl, 90% H_2O/10% D_2O, pH 6.5). ^{15}N and ^{15}N-^{13}C uniformly labelled samples were prepared replacing LB by M9 minimal medium containing $^{15}NH_4Cl$ and/or ^{13}C-glucose as the sole nitrogen and carbon sources following the same procedure as for the unlabeled sample. Protein samples homogeneity was checked by SDS–PAGE and protein concentration was determined by measuring the UV absorption at 280 nm, using the theoretical molar extinction coefficient of 8480 M^{-1} cm^{-1}.

2.2. NMR spectroscopy

NMR spectra were recorded on 0.5 mM unlabeled, 0.3 mM ^{15}N labeled or 0.3 mM $^{13}C/^{15}N$ double-labeled protein samples at 298 K on a Bruker Avance III 800 spectrometer equipped with a TXI 5 mm triple resonance probe. Backbone and side-chains resonances were assigned from the following spectra analysis: ^{15}N HSQC, ^{15}N HSQC TOCSY, obtained with the ^{15}N labeled sample and ^{13}C HSQC, HNCO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, HCCH-COSY and HCCH-TOCSY obtained with the $^{13}C/^{15}N$ double-labeled samples. Inter-proton distance restraints were derived from a homonuclear 2D NOESY (100 ms mixing-time) acquired using the unlabeled sample, a ^{15}N HSQC NOESY (100 ms mixing-time) acquired using the ^{15}N labeled sample and two ^{13}C HSQC NOESY (100 ms mixing-time) optimized for either aliphatic or aromatic detection, acquired using the $^{13}C/^{15}N$ labeled protein sample. Amide groups involved in hydrogen bonds were determined by a ^{15}N HSQC spectrum recorded immediately after solvent exchange from H_2O to D_2O by lyophilization/re-dissolution. Spectra were acquired and processed using TopSpin v3.1 (Bruker, Germany) and analyzed with Sparky v3.1 (T. D. Goddard and D. G. Kneller, University of California, San Francisco).

2.3. NMR structure calculation

The programs ARIA v1.2 (Linge et al., 2001) coupled to CNS v1.1 (Brünger et al., 1998) were used for automated NOE assignment and protein structure calculation by restraint simulated annealing, respectively. The process as implemented in ARIA consisted of 9 iterations of progressive assignment and structure calculations. In the last iteration, 200 structures were calculated and the 20 lowest-energy structures were refined in an explicit water box and taken as representative ensemble. As input data for the structure calculation, we used: (a) an almost complete ^{1}H, ^{15}N and ^{13}C resonance assignment list; (b) NOESY peak lists from four NOESY spectra, including manually assigned NOEs characteristic of helical secondary structure (H_N i–H_N i+1; Hα i–H_N i+2, i+3 and i+4 and Hα i–Hβ i+3) and characteristic of β-sheet secondary structure (long-range H_N–H_N, Hα–H_N and Hα–Hα); (c) phi and psi dihedral angles restraints, which were predicted by using the chemical shifts values in the program TALOS+ (Shen et al., 2009). Phi and psi errors estimates were also employed as predicted by the program; (d) hydrogen-bond restraints determined by a consensus of hydrogen/deuterium exchange experiment, CSI – derived secondary structure patterns (Wishart and Sykes, 1994) and manually assigned NOEs; (e) some unambiguous NOEs between aliphatic and aromatic side chains were also manually assigned and incorporated. The quality of the calculated structure ensemble was evaluated using MolMol (Koradi et al., 1996) and PROCHECK-NMR (Laskowski et al., 1996). Structural similarity search was performed using DALI (Holm and Sander, 1995), VAST (Gibrat et al., 1996) and PDBeFold (Krissinel and Henrick, 2004).

The structure atomic coordinates and restraints used in the calculation were deposited in the Protein Data Bank (PDB ID: 2MNI), and ^{1}H, ^{15}N, ^{13}C NMR chemical shift assignments (López-Castilla et al., 2014) were deposited in the BMRB database (BMRB ID: 19893).

2.4. ^{15}N relaxation rates and ^{1}H–^{15}N heteronuclear NOEs

^{15}N relaxation data were obtained from two series of 2D ^{1}H–^{15}N correlated spectra (1024 × 256 data points). Relaxation delays of 12, 52, 102, 202, 402, 902, 2002, and 4002 ms and 6, 10, 18, 34,

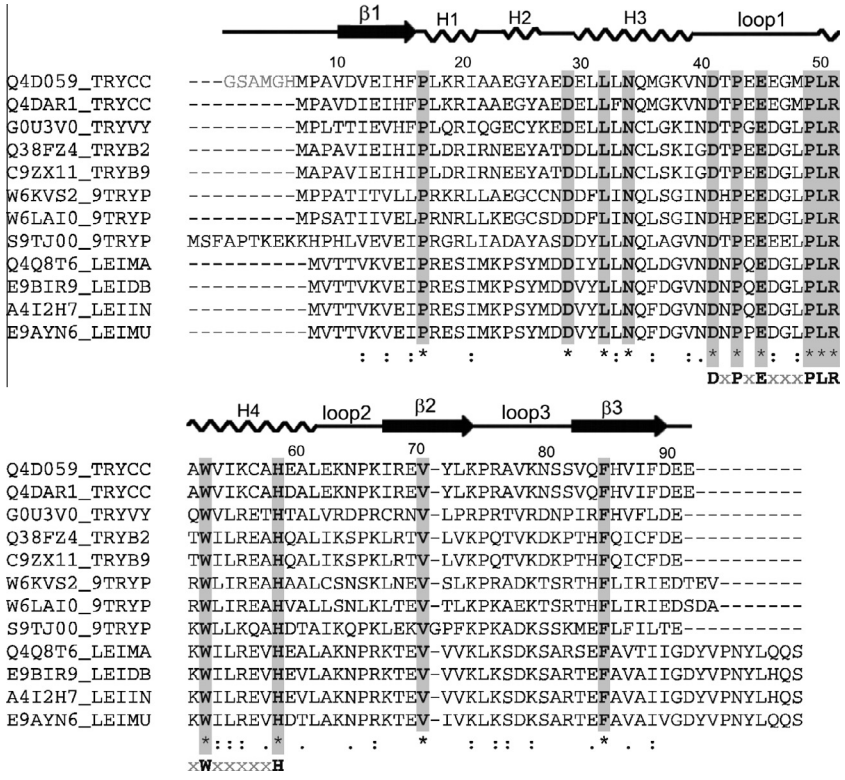

Fig.1. Sequence alignment of Q4D059 with homologues sequences. The protein sequences are identified by their UniProt code as follows: Q4D059_TRYCC and Q4DAR1_TRYCC from *Trypanosoma cruzi*, G0U3V0_TRYVY from *Trypanosoma vivax*, Q38FZ4_TRYB2 from *Trypanosoma brucei brucei*, C9ZX11_TRYB9 from *Trypanosoma brucei gambiense*, W6KVS2_9TRYP from *Phytomonas* spp. *isolate EM1*, W6LAI0_9TRYP from *Phytomonas* spp. *isolate Hart1*, S9TJ00_9TRYP from *Strigomonas culicis*, Q4Q8T6_LEIMA from *Leishmania major*, A4I2H7_LEIIN from *Leishmania infantum*, E9BIR9_LEIDB from *Leishmania donovani*, E9AYN6_LEIMU from *Leishmania mexicana*. Strictly conserved residues are highlighted by grey background and bold letters. Secondary structure elements, as determined experimentally, are depicted on top. A conserved sequential motif {D-x-P-x-E-x(3)-P-L-R-x-W-x(5)-H} is displayed below the trypanosomatid sequences alignment.

82, 162, 202, and 242 ms were used for T_1 and T_2, respectively. Recycling delay was 1.2 s and the number of scans was set to 8. Cross-peak intensities were fitted as implemented in SPARKY v.3.1 (T.D. Goddard and D.G. Kneller, University of California, San Francisco) to obtain T_1 and T_2 values that were plotted as relaxation rates $R_1 = 1/T_1$ and $R_2 = 1/T_2$. The error in the decay function fit was considered as the uncertainty in these cases.

Steady-state {1H}-15N NOEs were determined from two $^{1}H–^{15}N$ correlated spectra (1024 × 256 data points) according to the relation $NOE = I/I_{ref}$, where I is the cross-peak intensity in a 2D $^{1}H–^{15}N$ correlated spectrum with broadband ^{1}H pre-saturation and I_{ref} is the intensity without pre-saturation. Recycling delay was set to 2 s and the number of scans was 32. The uncertainty was set to 5% of the NOE value, estimated by sum of noise/signal ratio in both spectra. All relaxation experiments were performed on the ^{15}N labeled protein sample.

Global correlation time (τ_c) was estimated from the R_2/R_1 ratio assuming molecular isotropic tumbling, using the program Tensor2 (Dosset et al., 2000). Residues with NOE values below the average and/or undergoing conformational exchange were excluded for estimating τ_c. Relaxation data were analyzed for local mobility according to Lipari-Szabo model-free formalism using the program Tensor2 (Dosset et al., 2000).

3. Results and discussion

3.1. Primary sequence analysis

According to the OrthoMCL database (http://www.orthomcl.org/; release 10-May-2013), Q4D059 belongs to the OG5_154741 group that contains nine orthologs from the genomes of *Leishmania infantum*, *Leishmania major*, *Leishmania mexicana*, *Trypanosoma brucei gambiense*, *Trypanosoma brucei*, *T. cruzi* and *Trypanosoma vivax*. By searching the UniProt database using BLAST (Altschul et al., 1990), another three proteins from *Strigomonas culicis* and *Phytomonas sp.*, with significant sequence similarity (BLAST E-value ⩽ 2.0 E-12 and sequence identity ⩾ 39%) were found. All homologues (Fig. 1) are exclusively found in trypanosomatids. Since all proteins were annotated as hypothetical uncharacterized proteins of unknown function, functional inferences based on the primary sequence were not possible. The overall high-sequence identity among the twelve trypanosomatid proteins suggests that all of them may exhibit a similar fold and share a similar function on the parasites.

3.2. Secondary structure

The NMR structure in solution of Q4D059 was determined using 0.5 mM unlabeled, 0.3 mM ^{15}N uniformly labeled and a 0.3 mM ^{15}N and ^{13}C uniformly double-labeled protein samples. By means of standard NMR experiments, almost complete ^{1}H, ^{15}N and ^{13}C chemical shift assignments were obtained and the extension of the secondary structure elements were determined by a combination of secondary chemical shift analysis (López-Castilla et al., 2014), NOE pattern and $^{1}H/^{2}H$ exchange experiments. The resulting secondary structure elements are displayed on top of the primary sequences in Fig. 1.

Twelve unambiguous NOEs typical of β-sheet secondary structure, e.g. amide–amide, amide–Hα or Hα–Hα protons were manually assigned (black arrows in Fig. 2A), and allowed to define paring

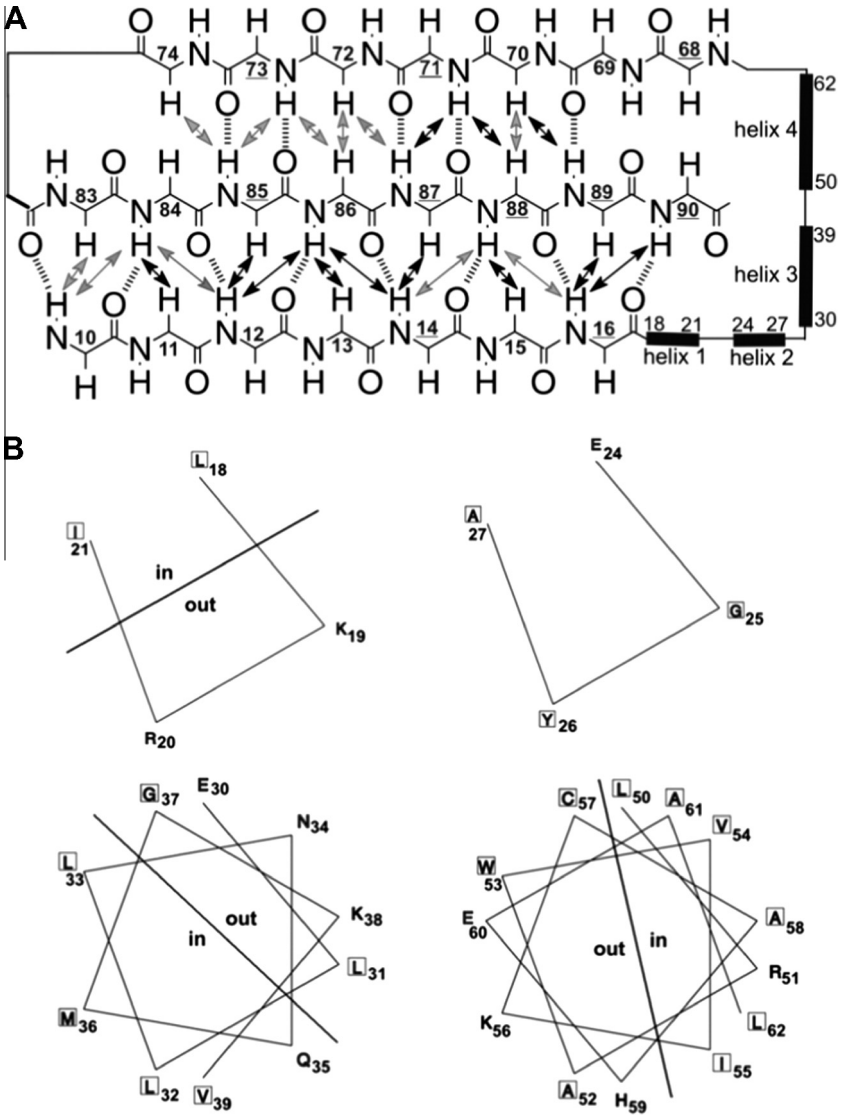

Fig.2. Secondary structure elements of Q4D059. (A) The backbone of residues in β-strands are represented and numbered (β1: 10–16, β2: 68–74 and β3: 83–90). Manually assigned NOEs are shown by arrows (black and gray arrows indicate unambiguously and ambiguously assigned NOEs, respectively), hydrogen bonds assigned on the basis of $H/^2H$ exchange experiments and NOEs pattern are shown by dotted lines. (B) The four helical structure elements are displayed in a wheel representation. Charged polar amino acids are not surrounded by squares.

and orientation of the β-strands predicted by chemical shift analysis, resulting in a parallel/anti-parallel three-stranded beta-sheet (Fig. 2A). Eleven of the expected NOEs between the β-strands were in principle ambiguous (grey arrows in Fig. 2A) but could be assigned based on the unambiguous NOEs (black arrows). Thirteen H-bonds were assigned in the β-sheet (dashed lines in Fig. 2A) and were used during the protein structure calculation. Nine of them were attributed based on the slow $^1H/^2H$ exchange observed on the corresponding amide protons (residues underlined in Fig. 2A) and the first four H-bonds between β1 and β3 were supported by the NOE pattern.

Additionally, twenty-six NOEs characteristics of helical secondary structure were also manually assigned, specifically amide–amide (H_N i–H_N i+1), Hα–amide (Hα i–H_N i+3) and (Hα i–Hβ i+3). In the calculation process up to eighty-two of these type of NOEs were automatically assigned, including (H_N i–H_N i+2), (Hα i–H_N i+2) and (Hα i–H_N i+4) (Fig. S1). Fig. 2B is a wheel representation of the protein helical segments. Most of them showed amphipathic features, suggesting that near half of the residue side-chains of each helix should be facing the structure core and the other part pointing to the protein surface.

3.3. Q4D059 tertiary structure

The NMR tridimensional structure of the protein Q4DO59 was calculated in a semi-automated fashion, using the methodology implemented in the program ARIA (Linge et al., 2001).

It was mainly based on NOE-derived H-H distance restraints, obtained from the analysis of 2D NOESY, ^{15}N HSQC NOESY and ^{13}C HSQC NOESY spectra. A small number of signals from the NOESY spectra were assigned manually, mainly those typical of secondary structure involving backbone atoms, Figs. 2A and S1. Some NOEs involving side-chain were also manually assigned, mainly between atoms with well-resolved resonances caused by ring-current shifts (e.g. I^{14}, L^{32}, V^{54}, L^{62}, L^{73}, I^{88} methyl groups) and aromatic side-chains (e.g. H^{15}, F^{16}, W^{53}, L^{62}, F^{85}, F^{89}). Nine manually assigned NOEs related to side chains interactions allowed us to roughly define the spatial orientation of the helical regions in relation to the β-sheet. The first three helical regions (residues 18–21, 24–27 and 30–39) showed NOEs mainly with residues in strand β1 (e.g. L^{32}Hδ1 – F^{16}Hε, M^{36}Hε – I^{14}Hγ2) and the helix H4 (residues 48–62) showed NOEs with residues located at strands β2 and β3 (e.g. L^{50}Hδ1 – F^{85}Hδ1, C^{57}Hβ – L^{73}Hδ1). In total

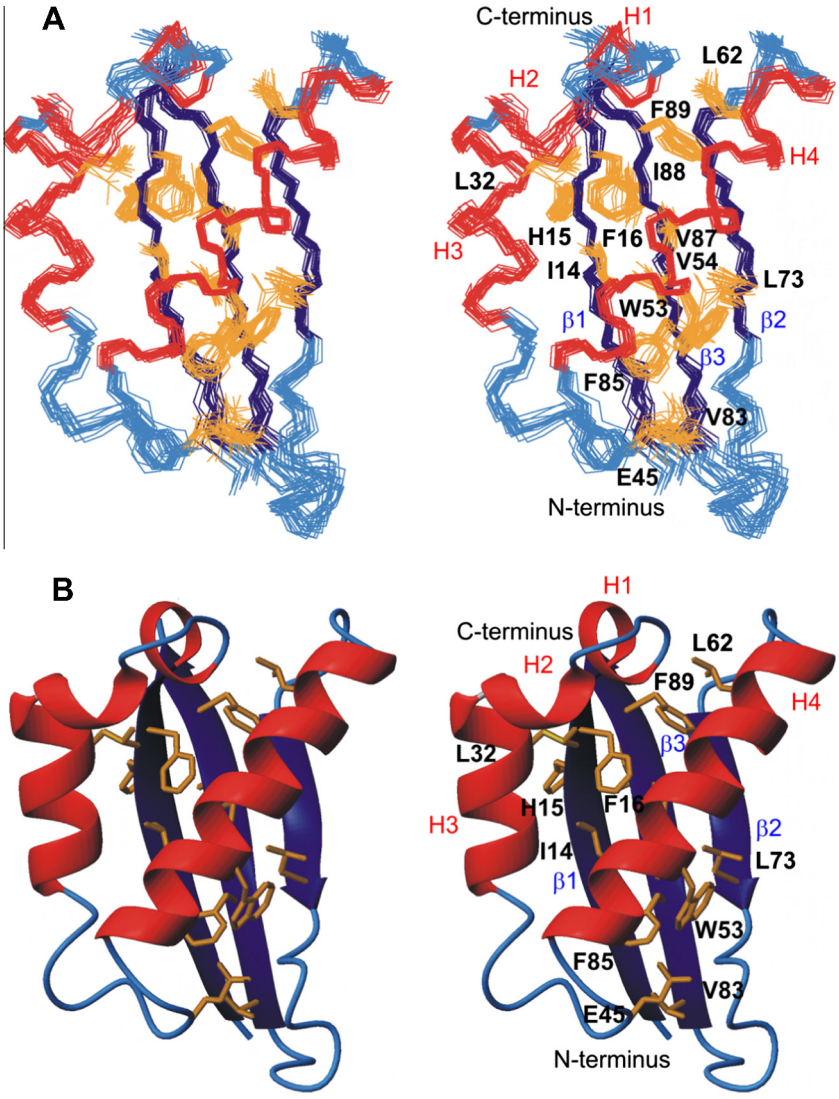

Fig.3. Solution NMR structure of Q4D059 from *T. cruzi*. (A) Ensemble of the 20 lowest energy structures, residues 7–92, in stereo view. Helical regions and β-strands are colored red and dark blue respectively, and loops are colored light blue. (B) Ribbon representation of Q4D059 lowest energy structure. Coloring code is the same as in A. Side chains of residues with large chemical shifts deviation from expected values, as well as spatially close aromatic side chains, are shown and labeled.

eighty-six NOE-base distance restraints were obtained manually. During the calculation process, most of the NOEs restraints were automatically assigned by ARIA, yielding a final number of 1924 restraints.

In preliminary protein structure calculations, additionally to thirteen assigned H-bonds in the β-sheet (Fig. 2A), we used twenty-five ambiguous H-bonds in the helical regions compatible to helices 3–10, α or π, and ninety-four TALOS-derived phi and psi angle restraints. In the final calculation, the ambiguous H-bond restraints previously used were substituted by unambiguous H-bonds typical of α helices in the regions 30–39 and 48–62, because these segments showed characteristics of α helices during the preliminary calculations.

The ensemble of 20 lowest-energy conformers out of 200 calculated is displayed in Fig. 3A and a ribbon representation of the lowest-energy structure is shown in Fig. 3B. The backbone root-mean-square deviation (rmsd) value of the ensemble to the average structure is 0.50 ± 0.05 Å and the protein structure is well defined by ca. 9 medium and long range NOEs per residue. Complete structure statistics are presented in Table 1.

Q4D059 adopts an α/β topology that includes a three-stranded β-sheet, two short helical regions (helices 1 and 2) and two α helices (helices 3 and 4). These secondary structure elements are arranged in the sequential order β1–helix H1–helix H2–helix H3–helix H4–β2–β3 (Fig. 3). Residues comprised in secondary structure elements are β1: V^{10} to F^{16}, helix H1: L^{18} to I^{21}, helix H2: E^{24} to A^{27}, helix H3: E^{30} to V^{39}, helix H4: M^{48} to L^{62}, β2: R^{69} to L^{73} and β3: V^{83} to D^{90}. In the β-sheet, β1 contacts β3 in parallel orientation while β3 is anti-parallel oriented to β2. One side of the β-sheet is mainly composed of non-polar residues contributing to the hydrophobic core of the structure and the outside is mostly comprised of amino acids with charged/polar side-chains. The helical regions are packed against the hydrophobic side of the β-sheet and they also showed a pattern of hydrophilic side chains exposed to the solvent and hydrophobic residues facing the protein core. Side chains of residues with large chemical shifts deviation from expected values are spatially close to aromatic residues on the structure (side-chains displayed in Fig. 3).

Helix H1 mostly shows NOEs and H-bonds typical of $helix_{3-10}$ and the short helix H2 adopts an irregular geometry. In proteins in general, $helices_{3-10}$ are three to five residues in length, in contrast to α helices that are usually longer, comprising ten to twelve residues (Richardson and Richardson, 1988). Furthermore, $helices_{3-10}$ are generally found at the N- or C-terminal of helices

Table 1
Q4D059 structure statistics.[a]

(A) Completeness of resonance assignments (%)[b]	
Backbone	94.2
Side chain	84.2
Aromatic	100
(B) Number of restraints	
NOE restraints	1794
Intra-residual	588
Sequential	396
Medium-range	273
Long-range	384
Ambiguous	153
Hydrogen bonds (two restraints each)	38
Dihedral angles (phi, psi)	94
Total	1964
(C) Residual experimental restraints violations	
rmsd NOEs (Å)	0.04 ± 0.006
rmsd dihedral angles (°)	0.99 ± 0.14
Average No. of distance violation > 0.5	1.1 ± 0.4
Average No. of dihedral angle violation > 5°	0.2 ± 0.4
(D) CNS potential energy (kcal mol^{-1})	
E_{total}	−3075 ± 92
E_{bonds}	25 ± 2
E_{angles}	139 ± 12
$E_{impropers}$	75 ± 8
$E_{dihedral}$	457 ± 8
E_{vdW}	−267 ± 21
E_{elec}	−3614 ± 81
E_{noe}	101 ± 12
E_{cdih}	5.7 ± 1.5
(E) Model quality	
rmsd backbone atoms to average structure (Å)	0.50 ± 0.05
rmsd heavy atoms to average structure (Å)	1.06 ± 0.09
rmsd bond lengths to ideal geometry (Å)	0.0042 ± 0.0001
rmsd bond angles to ideal geometry (°)	0.58 ± 0.02
(F) MolProbity Ramachandran statistics (%)[c,d]	
Most favored regions	87.4
Allowed regions	13.3
Disallowed regions	1.7
(G) Global quality scores (raw/Z score)[c,d]	
Verify3D	0.31/−2.41
ProsaII	0.56/−0.37
Procheck (phi–psi)	−0.55/−1.85
Procheck (all)	−0.58/−3.43
MolProbity clash score	28.41/−3.35
(H) Model contents	
Ordered residue ranges	9–91
Total No. of residues	92
BMRB accession number	19893
PDB ID code	2MNI

[a] As recommended by the wwPDB task force (Montelione et al., 2013).
[b] Computed using AVS (Moseley et al., 2004).
[c] Computed in the PSVS server (Bhattacharya et al., 2007).
[d] For residues 9–91.

α (Barlow and Thornton, 1988). The two short helices 1 and 2 of the protein Q4D059 are consistent in length with helices$_{3-10}$. The N-terminal segment and loop 3, connecting β2 and β3, are the less convergent regions of the protein (Fig. 3A) with a relative higher local rmsd value (Fig. S2). The worse convergence of these regions results from the absence of resonance assignments for the first eight amino acids, and the few number of NOEs per residue observed in the beginning of β1 and in loop 3. To check whether these regions are genuinely flexible, backbone relaxation experiments were performed.

3.4. Q4D059 Dynamics from NMR measurements

The internal backbone mobility of Q4D059 was investigated through ^{15}N relaxation parameters R_1, R_2 and {1H}-15 N steady-state NOE, which were determined for all assigned amide group resonances, Fig. 4. For most residues R_1, R_2 and NOE, near to average values were obtained (average $R_1 = 1.4\ s^{-1}$, $R_2 = 8.9\ s^{-1}$, NOE = 0.8), compatible with low internal mobility for most regions. The correlation time (τ_c = 5.2 ns) was estimated from the relation R_2/R_1 for residues in rigid regions, using the program Tensor2 (Dosset et al., 2000) and is compatible to a monomeric state for a 10 kDa globular protein.

Residues from A^9 to D^{11} at the beginning of β1, K^{38}, V^{39} at the end of helix H3, and D^{41} in loop 1, E^{92} in the C-terminal showed R_2 values lower than the average, indicating rapid motion in the ps–ns timescale for these residues. In contrast, G^{25} (helix H2), I^{68} (β2) and V^{78} (loop 3) exhibited R_2 values two standard deviation larger than the average value, characteristic of internal motion in the μs–ms timescale, Fig. 4. Lipari-Szabo model-free analysis was performed using the program Tensor2, Fig. 4. For most residues, values of S^2 between 0.8 and 0.95 were obtained; thus indicating the absence of fast motion in the N-H vectors. In contrast, residues from A^9 to D^{11} and V^{39} showed S^2 values below 0.8, describing the presence of thermal motion. An effective correlation time (τ_e) around 1 ns was necessary to fit the relaxation data in the case of A^9, V^{10}, D^{11} and K^{38}, indicating that these regions (begin of β1 and end of helix H3) are flexible and undergo fast movements in the ps–ns timescale. Moreover, for amino acids in the helix H2 (G^{25}), helix H3 (L^{33} and N^{34}), loop 1 (from E^{45} to G^{47}), loop 3 (A^{77}, V^{78}) and I^{68} were obtained exchange rates (R_{ex}) ranging from 2 to 7 s^{-1} (Fig. 4), highlighting the presence of conformational exchange process in the μs–ms time scale.

By comparing the convergence of Q4D059 structure ensemble (Figs. 3A and S2) and the dynamic data in Fig. 4, we conclude that the less convergent regions, i.e. N-terminus and begin of β1, helix H2, loop 3 are genuinely dynamic. Although a clear tendency of residues V^{10} and D^{11} to adopt a β-strand geometry have been seen by the chemical shift analysis (López-Castilla et al., 2014) and/or NOEs (Fig. 2A), the lower local convergence at the begin of β1 is compatible to the relaxation parameters measured and the fast ^{1}H/^{2}H exchange showed for the H_N of residues V^{10} to E^{13} (Fig. 2A). The same is true for helix H2 which show a helical tendency but has less convergent, irregular geometry and end of helix H3, loop1 and loop 3, which are less convergent. All these regions are characterized by relaxation parameters indicating local mobility.

3.5. NMR structure quality assessment

Table 1 contains several parameters recommended by wwPDB validation task force (Montelione et al., 2013) for assessment of model quality. Most values in Table 1 are in the acceptable range for NMR structures (e.g. MolProbity clash score and rms-deviation of bond lengths and angles from ideal geometry). However, a considerable number of residues, exclusively in loops 1 and 3 are in the disallowed region of the Ramachandran map. These residues are: N^{40} (in 4 of 20 models), P^{43} (1/20), E^{44} (1/20), E^{45} (5/20), G^{47} (11/20) in loop 3 and P^{75} (2/20), S^{81} (5/20) in loop 3. As the violations are less systematic, these values indicate a lower resolution of these regions as a consequence of the smaller number of NOEs assigned for those residues, reflected also in higher local rmsd (Fig. S2). Another factor that could contribute to the poorer structural quality of these regions is conformational averaging, considering that these loops presented distinguished dynamics in the ^{15}N relaxation studies.

3.6. Structural similarity to other proteins in the PDB

Comparison of the solution NMR structure of Q4D059 to structures in the PDB by DALI (Holm and Sander, 1995), PDBeFold

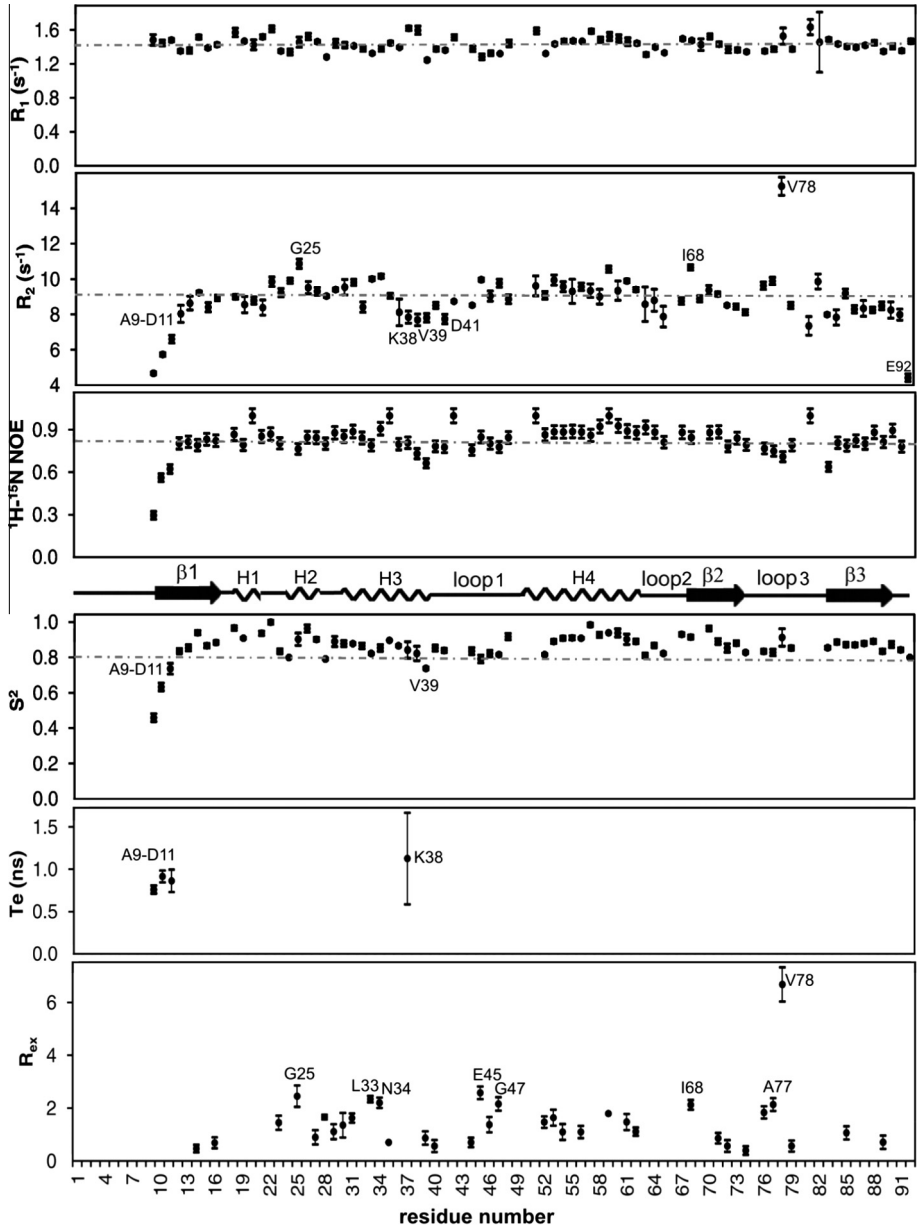

Fig.4. Q4D059 backbone dynamics from ^{15}N relaxation data. ^{15}N R_1, ^{15}N R_2 relaxation rates and {1H}-15N NOE, as well as the results from Lipari-Szabo model free analysis of the data, order parameter (S^2), effective correlation time (τ_e) and exchange rate (R_{ex}), are represented as a function of the residue number. The secondary structure elements of Q4D059 are represented schematically in the middle.

(Krissinel and Henrick, 2004) and VAST algorithms (Gibrat et al., 1996) revealed ca. 100 similar structures which suggest possible functional attributions for Q4D059. However, *Z*-scores obtained for the structural alignments were close to the threshold indicating low structural similarity or statistical significance ($2.1 \leqslant Z\text{-score} \leqslant 2.9$) and thus the structural similarity should be considered with care.

One of the top hits was the subdomain IIB of ATPase domain of a human heat shock protein 70 (HSP70, PDB ID: 3iuc-A, *Z*-score = 2.9, *Q*-score = 0.046, and sequence identity = 5%). The Q4D059 structure matches relatively well the subdomain IIB (rmsd = 3.1 Å over 59 aligned Cα atoms), although the strand β1 of Q4D059 is not present in the human protein (Fig. 5A and B). The subdomain IIB contains some of the residues responsible for ATP and ADP binding by interacting with the adenine moiety (Arakawa et al., 2011; Wisniewska et al., 2010). The amino acids involved in these interactions seem not to be conserved in structurally equivalent positions in Q4D059 what lower the probability of Q4DO59 binding adenine. The subdomain IIB also plays a role in the interaction with nucleotide exchange factors that stimulate the ADP-ATP exchange and the protein activity (Arakawa et al., 2010; Sondermann et al., 2001; Xu et al., 2008). Again, residues found to be important in the interaction with nucleotide exchange factors seem not to be conserved in Q4DO59.

Another structure found among the best hits is the N-terminal domain of the 50S ribosomal protein L11 (PDB ID: 2k3f, *Q*-score = 0.13, *Z*-score = 2.3, sequence identity = 13%). A superposition of Q4D059 and this structure is shown in Fig. 5A (rmsd = 4 Å over 56 aligned Cα atoms). Both domains differ mainly in the topology of the β-sheet (Fig. 5A and B). Ribosomal protein L11 is

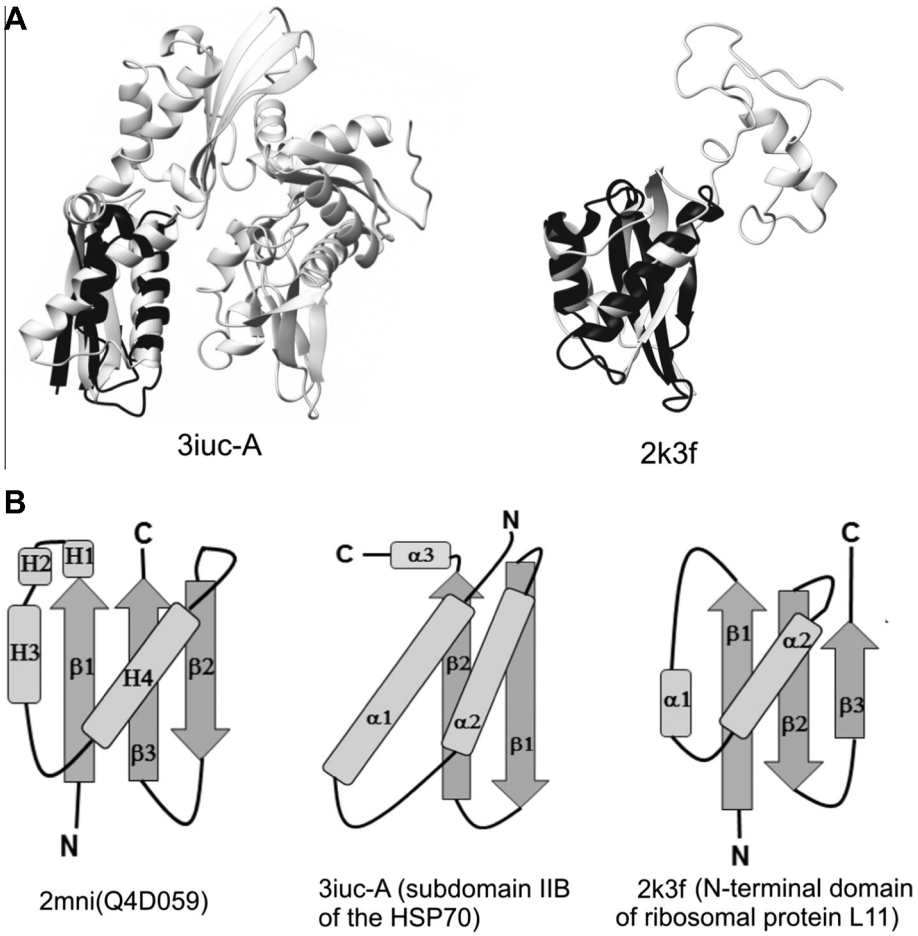

Fig.5. Structure comparisons of Q4D059 and structural homologues identified by DALI, VAST and PDBeFold. (A) Superposition of Q4D059 protein (dark-gray) with the ATPase domain of HSP70 (light-gray, PDB ID: 3iuc-A) and the ribosomal protein L11 (light-gray, PDB ID: 2k3f). (B) Topology diagrams for comparison of protein folds. Helical structures are represented by light grey cylinders and strands are shown by dark gray arrows.

a conserved protein containing two domains. The C-terminal domain makes major interactions with RNA while the N-terminal domain establishes only few specific contacts with the nucleic acid molecule. RNA-interacting residues in the N-terminal domain of ribosomal protein L11 include K^{10}, Q^{12}, Q^{30}, K^{71} (Wimberly et al., 1999) located at the exposed face of the β-sheet and adjacency. Among the residues found at the equivalent face in the Q4D059 structure, are two histidines and lysines/arginines that could grant some affinity to the negatively charged RNA in a pH dependent manner. The N-terminal domain of ribosomal protein L11 also participates in protein–protein interactions, e.g. with translation factors during protein synthesis (Agrawal et al., 2001; Stark et al., 2002; Valle et al., 2003; Van Dyke and Murgola, 2003). Structural studies supporting these findings are based on cryo-EM data and because of the resolution achieved, residues involved in the interactions could not be inferred (Agrawal et al., 2001; Petry et al., 2005; Stark et al., 2002; Valle et al., 2003; Datta et al., 2005; Connell et al., 2007; Harms et al., 2008; Anger et al., 2013).

Because of the low statistical significance obtained in such structural comparisons, experimental data is essential to further evidence the functional role of Q4D059.

Fig. 5B compares the differences in the secondary structure elements arrangement (number and/or topology of the β-strands and orientation and/or lengths of the helixes) among the similar structures and Q4D059 (Fig. 5B). The β-sheet of the subdomain IIB of the ATPase domain (3iuc-A) contains only two strands and they are longer than in Q4D059. Moreover, it has an extra helix (helix α3) at the C-terminal. In the case of the N-terminal domain of the ribosomal protein L11 (2k3f), the orientation of the β-strands (all antiparallel) is different from the strands in Q4D059, the helix α1 is shorter than helix H3 in the trypanosome protein and the two short N-terminal helical structures of Q4D059 (helix H1 and H2) are not present in the ribosomal protein L11 (Fig. 5B). Furthermore; no significant matches (SSAP > 80) were identified in the structural classification databases CATH (Greene et al., 2007) and SCOP (Murzin et al., 1995).

These observations led us to conclude that Q4D059 and the other two domains are to some extent similar but distinct in details. Thus, Q4D059 structure could be categorized as a new variant of the previous folds, belonging to the CATH classification 3.30, α/β 2-layer sandwich.

3.7. Conserved motif within the family

From the multiple sequence alignment of identified HPs (Fig. 1, sequence identity ranging: 39–97%), we observed a stretch of conserved residues: {D-x-P-x-E-x(3)-P-L-R-x-W-x(5)-H}, located from D^{41} to H^{59} (residue numbers according to Q4D059 sequence), that occurs exclusively in this family. Indeed, searching this motif in all proteins deposited in the UniProtKB/Swiss-Prot (release 2014_04: 544996 entries) and UniProtKB/TrEMBL (release 2014_04: 54958551 entries) databases, using PrositeScan (de Castro et al., 2006), only twelve hits in twelve sequences were found, which correspond to the same sequences in the alignment (Fig. 1). For the

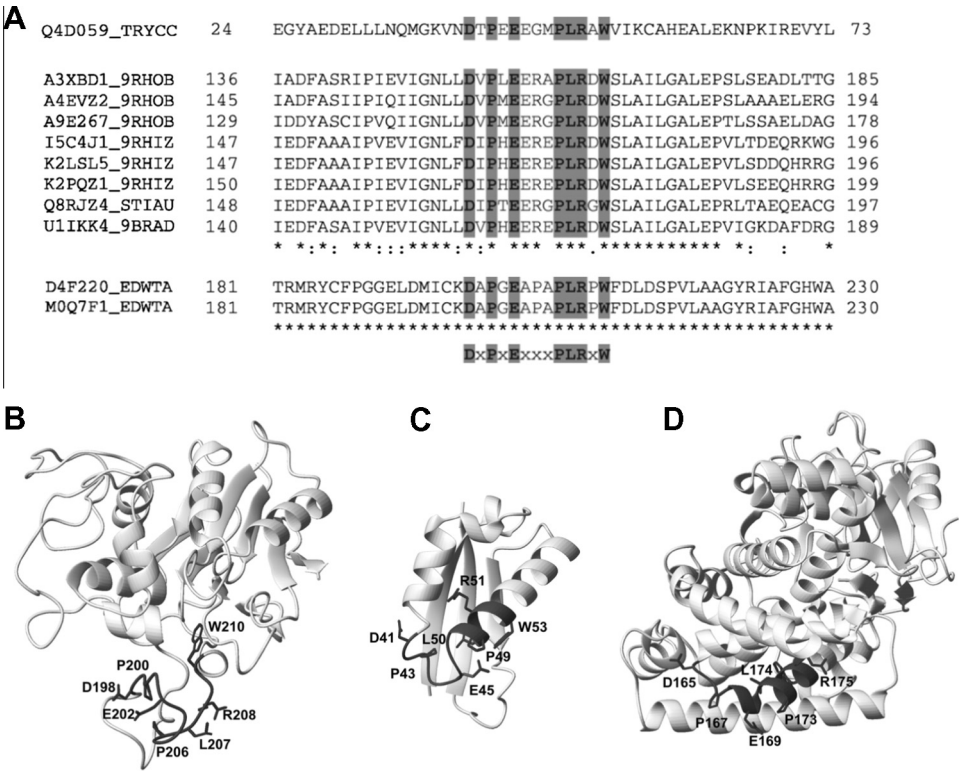

Fig.6. Conservation of the DxPxE/PLRxW motif from kinetoplastid proteins in sequences from cytochrome P450 family and bis (5′-nucleosyl)-tetraphosphatase. (A) A segment of protein sequences are shown and identified by their UniProt code as follows: Q4D059_TRYCC, HP from *Trypanosoma cruzi*; A3XBD1_9RHOB and A4EVZ2_9RHOB, cytochromes from *Roseobacter sp*; A9E267_9RHOB, cytochrome from *Oceanibulbus indolifex*; I5C4J1_9RHIZ, cytochrome from *Nitratireductor aquibiodomus*; K2LSL5_9RHIZ, cytochrome from *Nitratireductor pacificus*; K2PQZ1_9RHIZ, cytochrome from *Nitratireductor indicus*; Q8RJZ4_STIAU, cytochrome from *Stigmatella aurantiaca*; U1IKK4_9BRAD, cytochrome from *Bradyrhizobium sp*; D4F220_EDWTA and M0Q7F1_EDWTA, bis (5′-nucleosyl)-tetraphosphatases from *Edwardsiella tarda*. Strictly conserved residues are highlighted by grey background, bold letters. The motif {D-x-P-x-E-x(3)-P-L-R-x-W} is shown below the sequences. (B) Structural model of Q8RJZ4_STIAU, (C) structure of Q4D059 and (D) structural model of D4F220_EDWTA. In all cases, residues belonging to the DxPxE/PLRxW motif are shown in dark grey and side chains of strictly conserved residues are represented and labeled.

PrositeScan search performed the estimated number of expected random matches is 4.89 E-04 in ~100,000 sequences (50,000,000 residues) indicating low probability of finding these matches by chance. The sequence motif comprises eight of the fourteen strictly conserved residues (57%) in the protein family, and because most of them are in a loop (except L, R and W which are in a helix) and are exposed to the solvent (except L and W) the conservation of this motif could be of functional relevance.

Interestingly, using a motif one residue shorter, {D-x-P-x-E-x(3)-P-L-R-x-W}, in a PrositeScan search expands the matching sequences to ones non-homologous to Q4D059 (Fig. 6A). These include eight members of the large cytochrome P450 family and two representatives of bis(5′-nucleosyl)-tetraphosphatases from bacteria (Fig. 6A). According to PrositeScan (de Castro et al., 2006), {D-x-P-x-E-x(3)-P-L-R-x-W} has an approximate number of expected random matches of 2.20 E-02 in ~100,000 sequences, two orders of magnitude higher than the value obtained for the longer motif, yet indicating the low probability of finding these new sequences by chance. Indeed, performing a similar search, but against a randomized UniProtKB/Swiss-Prot database (reversed sequences), any hit is detected confirming this low chance probability.

The motif, {D-x-P-x-E-x(3)-P-L-R-x-W}, have not been previously annotated in Prosite and no function have been attributed for this stretch in the related cytochromes or phosphatases identified here. None of these proteins has their structures solved but some have been modeled by homology using proteins from yeasts, protozoa and bacteria as template (24–32% of sequence identity). In all models including the Q4D059 the motif sequence is a loop-helix segment, although the helix content and the spatial disposition of the conserved residues is different in each case (Fig. 6B–D).

4. Conclusions

The structure of Q4D059 is the first of a family of trypanosomatid proteins of unknown function. Because of the high sequence identity among members of the family, this structure can be used as template for structural modeling of other members.

Since only limited structural similarity was found to other proteins in the PDB (low statistical significance), the biological role of Q4D059 for the parasite remains to be experimentally investigated. The comparisons described here can guide the design of experiments. GST-pulldown and two-hybrid experiments could provide extra information about putative interaction partners, bringing insights into the protein function.

Structural and NMR data generated in this work will enable ligand-screening assays.

Acknowledgments

The authors acknowledge funding from the Brazilian agencies, CNPq, CAPES and FAPERJ.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2015.02.007.

References

Agrawal, R.K., Linde, J., Sengupta, J., Nierhaus, K.H., Frank, J., 2001. Localization of L11 protein on the ribosome and elucidation of its involvement in EF-G-dependent translocation. J. Mol. Biol. 311, 777–787.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., Horn, D., 2011. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res. 21, 915–924.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., Beckmann, R., 2013. Structures of the human and Drosophila 80S ribosome. Nature 497, 80–85.

Arakawa, A., Handa, N., Ohsawa, N., Shida, M., Kigawa, T., Hayashi, F., Shirouzu, M., Yokoyama, S., 2010. The C-terminal BAG domain of BAG5 induces conformational changes of the Hsp70 nucleotide-binding domain for ADP-ATP exchange. Struct. Lond. Engl. 1993 (18), 309–319.

Arakawa, A., Handa, N., Shirouzu, M., Yokoyama, S., 2011. Biochemical and structural studies on the high affinity of Hsp70 for ADP. Protein Sci. Publ. Protein Soc. 20, 1367–1379.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al., 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 38, D457–D462.

Atwood, J.A., Weatherly, D.B., Minning, T.A., Bundy, B., Cavola, C., Opperdoes, F.R., Orlando, R., Tarleton, R.L., 2005. The *Trypanosoma cruzi* proteome. Science 309, 473–476.

Barlow, D.J., Thornton, J.M., 1988. Helix geometry in proteins. J. Mol. Biol. 201, 601–619.

Bastos, I.M.D., Motta, F.N., Grellier, P., Santana, J.M., 2013. Parasite prolyl oligopeptidases and the challenge of designing chemotherapeuticals for Chagas disease, leishmaniasis and African trypanosomiasis. Curr. Med. Chem. 20, 3103–3115.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al., 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science 309, 416–422.

Bhattacharya, A., Tejero, R., Montelione, G.T., 2007. Evalutating protein structures determined by structural genomics consortia. Proteins 66, 778–795.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al., 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921.

Connell, S.R., Takemoto, C., Wilson, D.N., Wang, H., Murayama, K., Terada, T., Shirouzu, M., Rost, M., Schüler, M., Giesebrecht, J., Dabrowski, M., Mielke, T., Fucini, P., Yokoyama, S., Spahn, C.M., 2007. Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. Mol. Cell 9, 751–764.

Datta, P.P., Sharma, M.R., Qi, L., Frank, J., Agrawal, R.K., 2005. Interaction of the G′ domain of elongation factor G and the C-terminal domain of ribosomal protein L7/L12 during translocation as revealed by cryo-EM. Mol. Cell 20, 723–731.

De Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362–W365.

Dosset, P., Hus, J.C., Blackledge, M., Marion, D., 2000. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J. Biomol. NMR 16, 23–28.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.-N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., et al., 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309, 409–415.

Gibrat, J.F., Madej, T., Bryant, S.H., 1996. Surprising similarities in structure comparison. Curr. Opin. Struct. Biol. 6, 377–385.

Greene, L.H., Lewis, T.E., Addou, S., Cuff, A., Dallman, T., Dibley, M., Redfern, O., Pearl, F., Nambudiry, R., Reid, A., et al., 2007. The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Res. 35, D291–D297.

Harms, J.M., Wilson, D.N., Schluenzen, F., Connell, S.R., Stachelhaus, T., Zaborowska, Z., Spahn, C.M., Fucini, P., 2008. Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. Mol. Cell 30, 26–38.

Holm, L., Sander, C., 1995. Dali: a network tool for protein structure comparison. Trends Biochem. Sci. 20, 478–480.

Hotez, P.J., Bottazzi, M.E., Franco-Paredes, C., Ault, S.K., Periago, M.R., 2008. The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. PLoS Negl. Trop. Dis. 2, e300.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., Aert, R., et al., 2005. The genome of the kinetoplastid parasite, Leishmania major. Science 309, 436–442.

Koradi, R., Billeter, M., Wüthrich, K., 1996. MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. 14 (51–55), 29–32.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. 60, 2256–2268.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486.

Linge, J.P., O'Donoghue, S.I., Nilges, M., 2001. Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol. 339, 71–90.

López-Castilla, A., de Menezes, R.S., D' Andrea, E.D., Dos Santos, T.R., Pires, J.R., 2014. (1)H, (15)N and (13)C resonance assignments and secondary structure prediction of Q4D059, a conserved and kinetoplastid-specific hypothetical protein from *Trypanosoma cruzi*. Biomol. NMR Assign. http://dx.doi.org/10.1007/s12104-014-9565-z.

Minning, T.A., Weatherly, D.B., Atwood, J., Orlando, R., Tarleton, R.L., 2009. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. BMC Genomics 10, 370.

Montelione, G.T., Nilges, M., Bax, A., Güntert, P., Herrmann, T., Richardson, J.S., Schwieters, C.D., Vranken, W.F., Vuister, G.W., Wishart, D.S., Berman, H.M., Kleywegt, G.J., Markley, J.L., 2013. Recommendations of the wwPDB NMR Validation Task Force. Structure 21, 1563–1570.

Moseley, H.N., Sahota, G., Montelione, G.T., 2004. Assignment validation software suite for the presentation of protein resonance assignment data. J. Biomol. NMR 28, 341–355.

Muñoz-Saravia, S.G., Haberland, A., Wallukat, G., Schimke, I., 2012. Chronic Chagas' heart disease: a disease on its way to becoming a worldwide health problem: epidemiology, etiopathology, treatment, pathogenesis and laboratory medicine. Heart Fail. Rev. 17, 45–64.

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.

Petry, S., Brodersen, D.E., Murphy, F.V., Dunham, C.M., Selmer, M., Tarry, M.J., Kelley, A.C., Ramakrishnan, V., 2005. Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. Cell 123, 1255–1266.

Rassi, A., Rassi, A., Marin-Neto, J.A., 2010. Chagas disease. Lancet 375, 1388–1402.

Richardson, J.S., Richardson, D.C., 1988. Amino acid preferences for specific locations at the ends of alpha helices. Science 240, 1648–1652.

Shen, Y., Delaglio, F., Cornilescu, G., Bax, A., 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR 44, 213–223.

Sondermann, H., Scheufler, C., Schneider, C., Hohfeld, J., Hartl, F.U., Moarefi, I., 2001. Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. Science 291, 1553–1557.

Stark, H., Rodnina, M.V., Wieden, H.J., Zemlin, F., Wintermeyer, W., van Heel, M., 2002. Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. Nat. Struct. Biol. 9, 849–854.

Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R.E., McKerrow, J., Reed, S., Tarleton, R., 2008. Kinetoplastids: related protozoan pathogens, different diseases. J. Clin. Invest. 118, 1301–1310.

Valle, M., Zavialov, A., Li, W., Stagg, S.M., Sengupta, J., Nielsen, R.C., Nissen, P., Harvey, S.C., Ehrenberg, M., Frank, J., 2003. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. Nat. Struct. Biol. 10, 899–906.

Van Dyke, N., Murgola, E.J., 2003. Site of functional interaction of release factor 1 with the ribosome. J. Mol. Biol. 330, 9–13.

Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W., Ramakrishnan, V., 1999. A detailed view of a ribosomal active site: the structure of the L11-RNA complex. Cell 97, 491–502.

Wishart, D.S., Sykes, B.D., 1994. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J. Biomol. NMR 4, 171–180.

Wisniewska, M., Karlberg, T., Lehtio, L., Johansson, I., Kotenyova, T., Moche, M., Schuler, H., 2010. Crystal structures of the ATPase domains of four human Hsp70 isoforms: HSPA1L/Hsp70-hom, HSPA2/Hsp70-2, HSPA6/Hsp70B′, and HSPA5/BiP/GRP78. PLoS One 5, e8625.

Xu, Z., Page, R.C., Gomes, M.M., Kohli, E., Nix, J.C., Herr, A.B., Patterson, C., Misra, S., 2008. Structural basis of nucleotide exchange and client binding by the Hsp70 cochaperone Bag2. Nat. Struct. Mol. Biol. 15, 1309–1317.